Discrete Mathematics 25 (1979) 285–287

NOTE

THE TAIL OF THE HYPERGEOMETRIC DISTRIBUTION

V. CHVÁTAL
Department d'Informatique et Recherche Opérationnelle, Université de Montréal, CP612J, Montréal HEC 357, P.Q., Canada

Received 6 March 1978

Investigating random discrete structures, one often needs upper bounds on the tail of the hypergeometric distribution,

$$H(M, N, n, k) = \sum_{i=k}^{n} \binom{M}{i}\binom{N-M}{n-i}\binom{N}{n}^{-1}$$

with $k=(p+t)n$ for $p=M/N$ and some $t \geqslant 0$. A particularly useful bound,

$$H(M, N, n, k) \leqslant \left(\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{1-p}{1-p-t}\right)^{1-p-t}\right)^{n} \tag{1}$$

is included as a special case in a powerful result of Hoeffding [1]. Converting (1) into more elegant but weaker bounds is a routine matter. For example, setting

$$f(t) = (p+t)\log\frac{p}{p+t} + (1-p-t)\log\frac{1-p}{1-p-t}$$

we have

$$f'(t) = \log\frac{p(1-p-t)}{(p+t)(1-p)}$$

and

$$f''(t) = \frac{-1}{(p+t)(1-p-t)} \leqslant -4$$

whenever $0 \leqslant t \leqslant 1-p$. Since $f(0) = f'(0) = 0$, we conclude that $f(t) \leqslant -2t^2$ and so

$$H(M, N, n, k) \leqslant e^{-2t^2 n}.$$

The purpose of this note is to present a simple proof of (1). First of all, let us verify that

$$\binom{N}{n}^{-1}\sum_{i=j}^{n}\binom{M}{i}\binom{N-M}{n-i}\binom{i}{j} \leqslant \binom{n}{j}\left(\frac{M}{N}\right)^{j}. \tag{2}$$

Since

$$\sum_{i=j}^{n}\binom{M}{i}\binom{N-M}{n-i}\binom{i}{j}=\binom{M}{j}\sum_{i=j}^{n}\binom{M-j}{i-j}\binom{N-M}{n-i}=\binom{M}{j}\binom{N-j}{n-j},$$

the inequality (2) may be written as

$$\binom{N}{n}^{-1}\binom{M}{j}\binom{N-j}{n-j}\leqslant\binom{n}{j}\left(\frac{M}{N}\right)^{j}$$

which further simplifies into

$$\binom{M}{j}\binom{N}{j}^{-1}\leqslant\left(\frac{M}{N}\right)^{j}.$$

The last inequality is a direct consequence of $M\leqslant N$, and so (2) follows.

Now we may observe that for every $x\geqslant 1$ we have

$$\begin{aligned}\sum_{i=0}^{n}\binom{M}{i}\binom{N-M}{n-i}\binom{N}{n}^{-1}x^{i}&=\sum_{i=0}^{n}\binom{M}{i}\binom{N-M}{n-i}\binom{N}{n}^{-1}\sum_{j=0}^{i}\binom{i}{j}(x-1)^{j}\\&=\sum_{j=0}^{n}\left(\binom{N}{n}^{-1}\sum_{i=j}^{n}\binom{M}{i}\binom{N-M}{n-i}\binom{i}{j}\right)(x-1)^{j}\leqslant\sum_{j=0}^{n}\binom{n}{j}\left(\frac{M}{N}\right)^{j}(x-1)^{j}\\&=\left(1+\frac{(x-1)M}{N}\right)^{n}.\end{aligned}$$

Combining this observation with the fact that

$$H(M,N,n,k)\leqslant\sum_{i=0}^{n}\binom{M}{i}\binom{N-M}{n-i}\binom{N}{n}^{-1}x^{i-k}$$

we conclude that

$$H(M,N,n,k)\leqslant x^{-k}\left(1+\frac{(x-1)M}{N}\right)^{n}=(x^{-(p+t)}(1+(x-1)p))^{n}$$

for all $x\geqslant 1$. Choosing

$$x=\frac{(1-p)(p+t)}{p(1-p-t)}$$

we obtain the desired result (1).

Incidentally, an analogous bound on the tail of the binomial distribution,

$$B(p,n,k)=\sum_{i=k}^{n}\binom{n}{i}p^{i}(1-p)^{n-i}$$

is even easier to establish: we have

$$B(p,n,k)\leqslant\sum_{i=0}^{n}\binom{n}{i}p^{i}(1-p)^{n-i}x^{i-k}=x^{-k}(1+(x-1)p)^{n}$$

for all $x \geqslant 1$, and so

$$B(p,n,k) \leqslant \left(\left(\frac{p}{p+t} \right)^{p+t} \left(\frac{1-p}{1-p-t} \right)^{1-p-t} \right)^n \leqslant e^{-2t^2 n}$$

as above. (Apparently this argument has been used first by S.N. Bernstein; for more references, the reader is directed to [1].)

Acknowledgment

I am grateful to Persi Diaconis for bringing Hoeffding's paper to my attention.

Reference

[1] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Statist. Assoc. 58 (1963) 13–30.